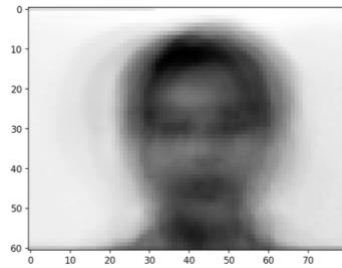


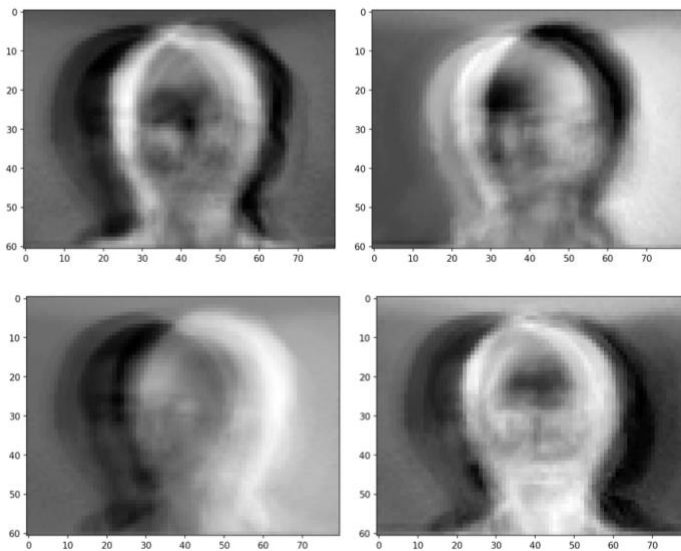
## 2 Programming Part

(a)

Mean

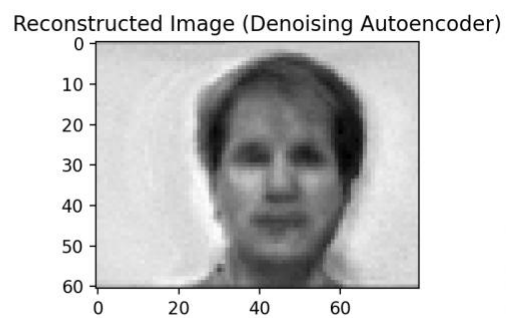
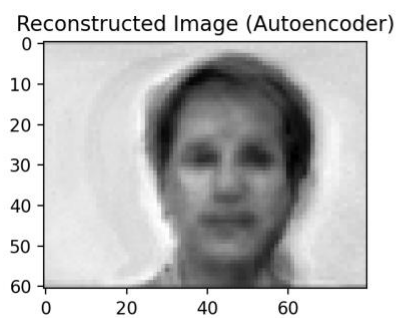
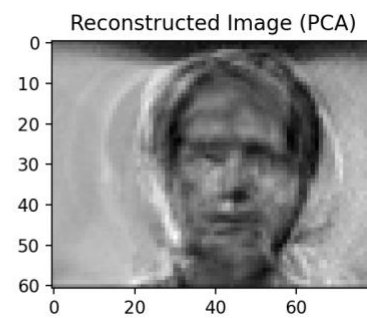
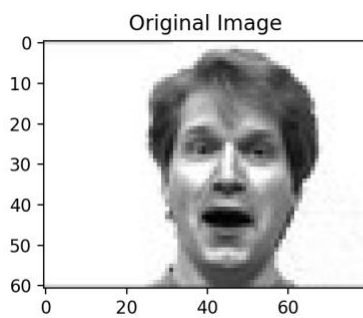


Top 4:

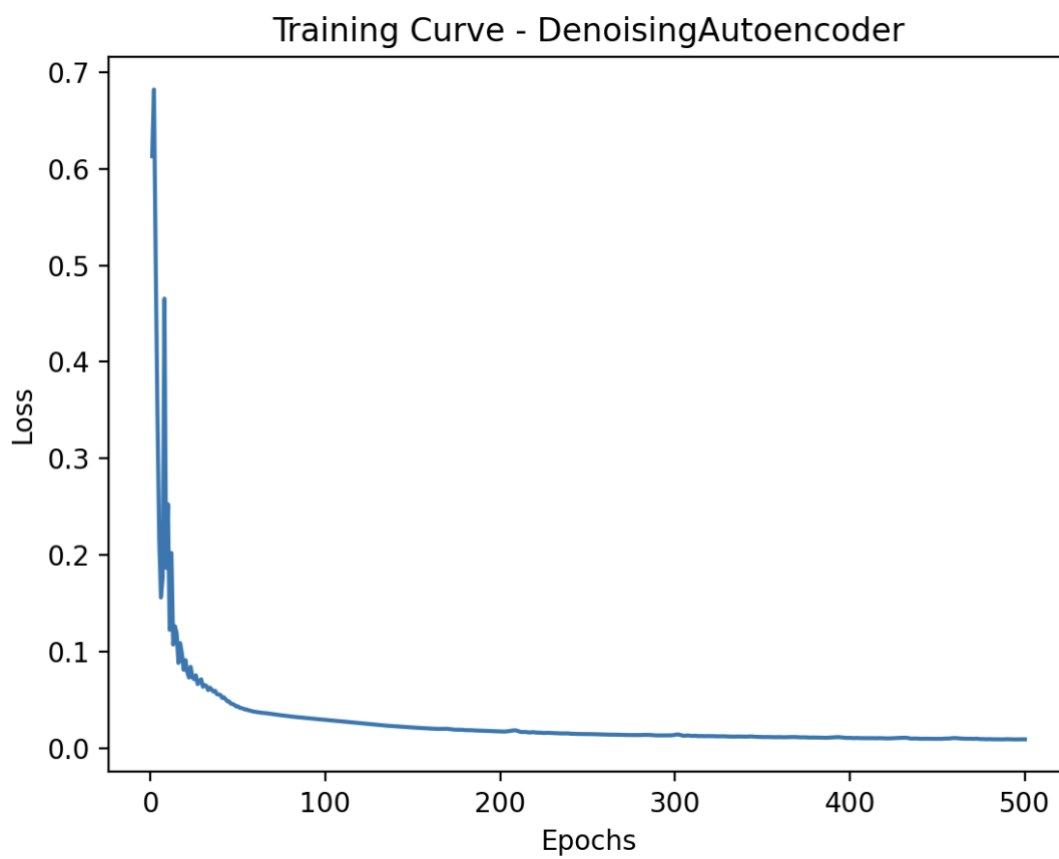
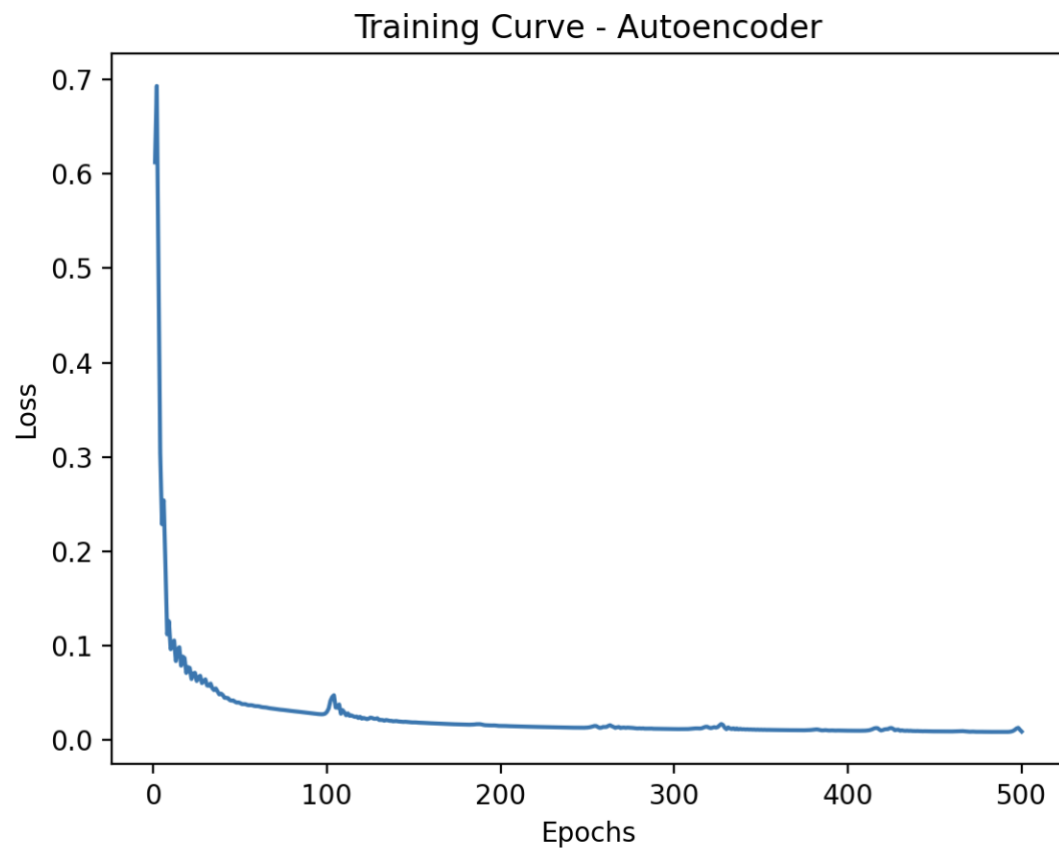


(b)

MSE: PCA = 0.0107, Autoencoder = 0.0148, Denoising Autoencoder = 0.0138



(c)



(d)

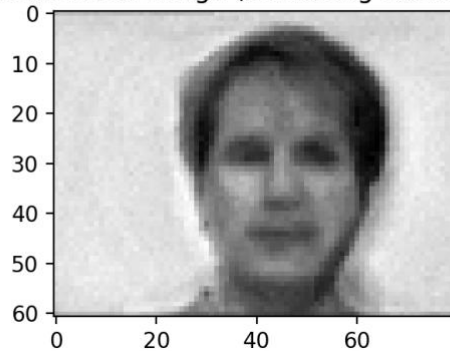
**Shallower:**

Encoder: Linear(4880, 488) -> Linear(488, 244) -> Linear(244, 122)

Decoder: Linear(122, 244) -> Linear(244, 488) -> Linear(488, 4880)

MSE: 0.0137

Reconstructed Image (Denoising Autoencoder)



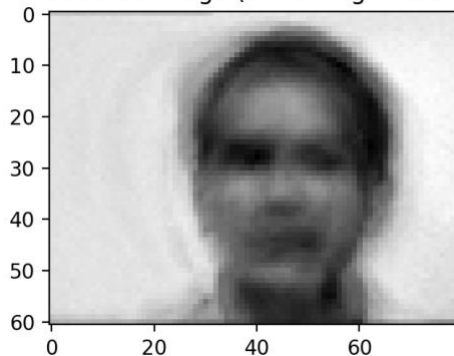
**Deeper:**

Encoder: Linear(4880, 1952) -> Linear(1952, 488) -> Linear(488, 244) -> Linear(244, 122)

Decoder: Linear(122, 244) -> Linear(244, 488) -> Linear(488, 1952) -> Linear(1952, 4880)

MSE: 0.0191

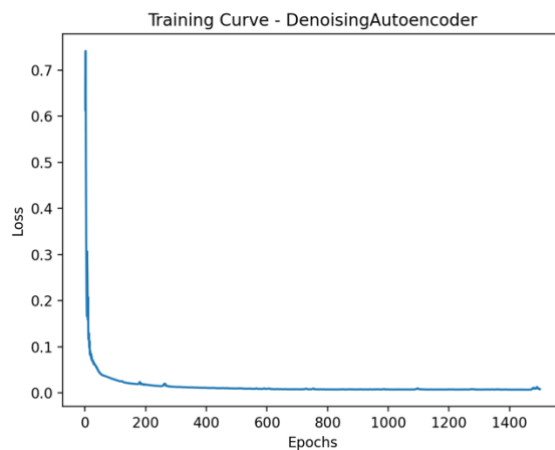
Reconstructed Image (Denoising Autoencoder)



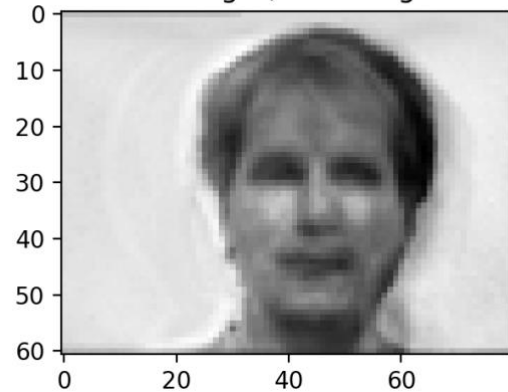
We can see that deeper model does not perform better in this case. The reconstruction error for shallower model is lower.

(e)

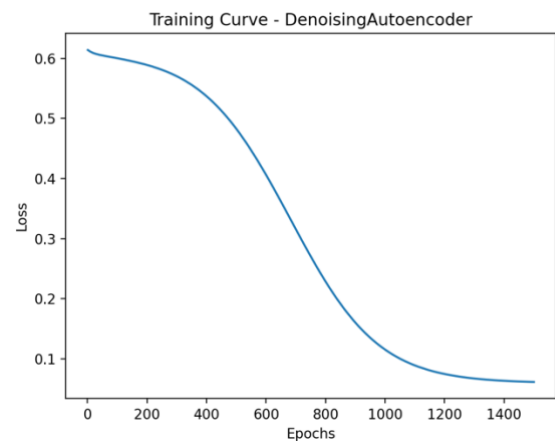
Adam:



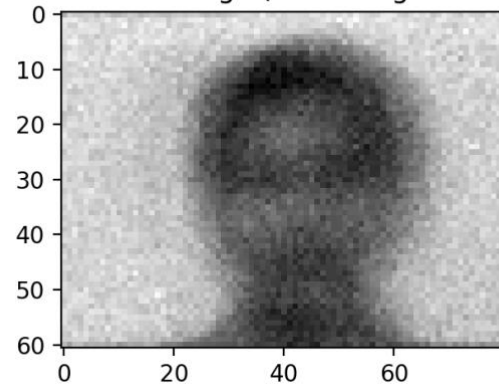
Reconstructed Image (Denoising Autoencoder)



SGD:



Reconstructed Image (Denoising Autoencoder)



From the training curves above, we can see that Adam has faster convergence speed and is able to find better local minimum; the reconstructed image also tells us that Adam(mse = 0.0132) performs better than SGD(mse = 0.0374).